

Datenbanken der Historischen Statistik im ZHSF

Metz, Rainer; Berg, Alexander

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Metz, R., & Berg, A. (2004). Datenbanken der Historischen Statistik im ZHSF. *Historical Social Research*, 29(2), 160-171. <https://doi.org/10.12759/hsr.29.2004.2.160-171>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Datenbanken der Historischen Statistik im ZHSF

Rainer Metz & Alexander Berg^{*}

Abstract: The availability of historical statistical data of high quality in machine-readable form for Germany is in a lamentable condition. The existing data are scattered over numerous publications. They are only partially available in machine-readable form and only a small part of them is stored in public archives. But not only that most of the existing machine-readable data are not archived, for Germany we also miss a statistical handbook for the history of economy, society and state - in printed as well as in electronic form - which presents the relevant figures in a systematic way. The online data bank HISTAT tries to offer the historical statistical data to the scientific community. This data bank does not only present the numerous figures and time series, it also documents the various annotations which help to understand the figures and the sources that have been the basis for creating the figures by the researcher, and, last but not least, it contains the literature which was consulted when editing the data. The final goal of this data bank is to present the data in a systematic way. Like the Statistical Yearbook, the online data bank must offer the categorized time series for different regions and branches, together with all information that are necessary to understand and evaluate historically the meaning of these statistical data.

^{*} Address all communications to: Rainer Metz, Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln, Abt. ZHSF, Liliencronstr. 6, 50931 Köln, Email: metz@za.uni-koeln.de; Alexander Berg, Fa. data-quest * Suchi&Berg GmbH, Geismar Landstr. 17b, 37083 Göttingen, Email: berg@data-quest.de.

Der Beitrag stellt die erweiterte Fassung eines Vortrages dar, der im Rahmen der Tagung "hist 2003: Geschichte und neue Medien" am 10. April 2003 an der Humboldt-Universität Berlin gehalten wurde.

I

Eine „Historische Statistik“ von Deutschland ist ein Desiderat der Forschung. Die umfangreichen Datenbestände, die von der Forschung aufbereitet und zusammengestellt wurden, sind zwar in zahlreichen, thematisch orientierten Bänden publiziert, ein Statistisches Handbuch, das modernen Ansprüchen genügen würde, fehlt jedoch. Zudem sind die zugrunde liegenden elektronischen Daten entweder überhaupt nicht, oder nur unter erschwerten Bedingungen verfügbar. Hier setzen die Bemühungen des Zentrums für Historische Sozialforschung (ZHSF) an und zwar in drei Stufen. Es wird erstens versucht, die Daten der „Historischen Statistik“, sofern sie noch maschinenlesbar vorliegen, zu archivieren und damit langfristig zu sichern. Zweitens werden bereits am ZHSF archivierte Daten in moderne Datenformate überführt, um so ihre schnelle und unkomplizierte Verwendung für Sekundäranalysen mit moderner Statistiksoftware zu gewährleisten. Schließlich sollen drittens die Daten unterschiedlicher Studien systematisch verknüpft werden. Ziel ist ein Handbuch zur „Historischen Statistik“ Deutschlands in elektronischer Form. Bevor wir uns mit diesen Vorhaben im Abschnitt III im Detail auseinandersetzen, soll im nächsten Abschnitt (II) kurz das ZHSF vorgestellt werden. In Abschnitt IV wird die EDV-technische Umsetzung der einzelnen Arbeitsschritte dargestellt.

II

Das ZHSF wurde 1987 auf Empfehlung des Wissenschaftsrates in das Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln (ZA) eingegliedert. Das ZA ist ein Institut der Gesellschaft sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS), die wiederum Mitglied der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL) ist.

Das ZHSF will im Grenzbereich zwischen Sozialwissenschaften und Geschichte infrastrukturelle Service- und Forschungsleistungen erbringen, die für eine qualitativ bessere und effizientere empirisch-methodische Forschung erforderlich sind. Eine der wichtigsten Aufgaben des ZHSF ist die Archivierung, Aufbereitung und Dokumentation maschinenlesbarer Forschungsdaten, die im Rahmen historischer Studien entstanden sind. Damit ist sichergestellt, dass dieses EDV-Material, das häufig unter großem finanziellen Aufwand aus den Archiven erhoben und aufbereitet wurde, für die wissenschaftliche Nutzung auch langfristig zur Verfügung steht. Neben der Archivierung der Daten wird vom ZHSF eine Dokumentation erstellt, die Entstehungszusammenhang und Inhalt der Studie dokumentiert. Damit ist eine problemlose spätere Verwendung dieser Daten gewährleistet. Im ZHSF sind gegenwärtig annähernd 160 historische Studien (darunter auch sehr umfangreiche) archiviert, von

denen ca. 140 Studien über eine standardisierte Kurzbeschreibung recherchierbar sind (www.gesis.org/Datenservice/Suche/Daten/index.htm). Neben Daten zur kollektiven Biographik, neben prozess-produzierten Daten und Wahldaten bilden Daten der „Historischen Statistik“ einen Schwerpunkt der Datenkollektion des ZHSF.

III

Was versteht man unter Daten der „Historischen Statistik“? Wie grenzen sich diese von den Daten der amtlichen Statistik einerseits und andererseits von historisch-quantitativen Daten schlechthin ab? Geht man dieser Frage nach, stellt man schnell fest, dass der Begriff der „Historischen Statistik“ nirgendwo eindeutig definiert ist und in unterschiedlichen Kontexten mit verschiedenen Inhalten belegt wird. Ganz allgemein kann man sagen, dass die „Historische Statistik“ quantitative Daten zur demographischen, wirtschaftlichen, sozialen, rechtlichen, politischen und kulturellen Entwicklung eines Landes vermittelt und dabei unterschiedliche, sowohl geographische wie auch institutionelle Ebenen berücksichtigt. Allerdings kann man „quantitative“ Daten und Daten der „Historischen Statistik“ nicht generell gleichsetzen, letztere sind vielmehr eine Teilmenge der ersteren. Man sollte nämlich nur dann von Daten der „Historischen Statistik“ sprechen, wenn es sich um quantitative Daten handelt, die über die Ausprägungen von inhaltlich genau definierten Begriffen informieren, die dann in Variablen operationalisiert werden, und die in historischen Untersuchungen aus zeitgenössischen amtlichen oder privaten Quellen systematisch erhoben, oder aus solchen Quellen geschätzt wurden. Damit ergibt sich eine wichtige Abgrenzung zu den Daten der amtlichen Statistik – diese werden nämlich nicht im Rahmen historischer Untersuchungen quasi im Nachhinein, sondern immer aktuell aus der jeweiligen Gegenwart erhoben.¹

Eine weitere Einschränkung hat der Begriff der „Historischen Statistik“ durch das Schwerpunktprogramm der Deutschen Forschungsgemeinschaft (DFG) erfahren, das seine primäre Aufgabe darin sah, vorwiegend Langzeitreihen zu wirtschaftlichen und sozialen Variablen zu ermitteln.² Die „Historische Statistik“ von Deutschland wurde dabei konzipiert als ein zeitlich möglichst weit zurückgreifendes, regional tief gegliedertes und thematisch breit angelegtes Dienstleistungsunternehmen, um für zukünftige Forschungen unterschiedlichster Art quellenkritisch kommentierte Daten aus den Bereichen Bevölke-

¹ Vgl. Ehling, Manfred: Historische Statistik – Probleme und Perspektiven der internationalen Zusammenarbeit. In: Statistisches Bundesamt (Hg.): Wirtschaft und Statistik 1996/7, S. 413-421.

² Vgl. den umfassenden Überblick in Kunz, Andreas: Historische Statistik von Deutschland. Ein Forschungsschwerpunkt der Deutschen Forschungsgemeinschaft 1981-1991. In: *Historical Social Research* 22/2, 1997, S. 236-249 mit umfangreichen Literaturangaben.

rung, Wirtschaft, Gesellschaft und Staat bereitzustellen. Besonderer Wert wurde darauf gelegt, die ermittelten Daten in ihrem Entstehungs- und Bedeutungszusammenhang kritisch zu erläutern und zu kommentieren. Ein Vorhaben, das natürlich umso bedeutsamer wird, je weiter die Daten in die Geschichte zurückreichen.

Sieht man einmal davon ab, dass die Definition dessen, was man unter einer „Langzeitreihe“ verstehen soll, immer problematisch sein wird, erscheint uns eine Beschränkung der „Historischen Statistik“ auf lange Zeitreihen wenig sinnvoll, da damit wichtige Datenkollektionen, die kurze Zeiträume umfassen, von vorne herein ausgeklammert würden. In unserem Zusammenhang sollen deshalb unter Daten der „Historischen Statistik“ Zeitreihendaten verstanden werden, die den oben genannten Bedingungen historischer Statistikdaten genügen. Für die historische Sozial- und Wirtschaftsforschung sind zeitreihenfähige Daten seit jeher von besonderer Bedeutung. Erst sie ermöglichen es, Zustände und deren kurz- und langfristige Veränderungen systematisch darzustellen und statistisch zu analysieren. Daten einer so definierten „Historischen Statistik“ sind für Deutschland zahlreich publiziert. So hat allein der DFG-Schwerpunkt bis heute zu 26 Statistikbänden mit unterschiedlichen Schwerpunkten geführt. Eine Vielzahl thematisch breit gestreuter Daten der „Historischen Statistik“ findet sich darüber hinaus in Datenhandbüchern, Monographien und Aufsätzen, die entweder im Rahmen größerer Forschungsprojekte oder im Kontext von wissenschaftlichen Qualifikationsarbeiten entstanden sind.³

Leider ist nur ein kleiner Teil dieser Daten in maschinenlesbarer Form am ZHSF archiviert. Dies hat mehrere Gründe: Erstens sind zahlreiche Daten, deren Generierung schon länger zurückliegt, überhaupt nicht mehr in maschinenlesbarer Form vorhanden. Zweitens kommt es vor, dass sich die vorhandenen Daten nicht mehr verarbeiten lassen, da die Datenträger defekt bzw. nicht mehr lesbar sind. Ein dritter Grund ist in der Tatsache zu sehen, dass es nicht selten auch an der Bereitschaft seitens der Primärforscher fehlt, ihre Daten an das ZHSF zur Archivierung und Aufbereitung weiterzugeben. Ein wichtiges Ziel des ZHSF ist es deshalb, alle Daten der „Historischen Statistik“, sofern sie bei den Primärforschern noch maschinenlesbar vorliegen, zu archivieren und damit langfristig zu sichern. Entsprechende Bemühungen haben hier zu ersten Teilerfolgen geführt. Ein weiteres Ziel des ZHSF ist es, bereits archivierte Daten, sofern sie in alten Formaten gespeichert sind, entweder in aktuelle Datenformate (Excel, SPSS) oder in Datenbanken zu überführen.

Eine besondere Bedeutung haben in diesem Zusammenhang Daten aus dem Projektverbund „Deutsche Bildungsstatistik 1800-1945“, der von 1977 bis 1981 von der DFG gefördert wurde. Die Projekte hatten das Ziel, auf breiter empirisch-statistischer Grundlage den langfristigen Strukturwandel des Bil-

³ Vgl. auch Kaufhold, Karl Heinrich: Neuere Quellen und Veröffentlichungen zur historischen Statistik von Deutschland. In: Historische Zeitschrift 262, 1996, S. 127-136.

dungswesens in Deutschland zu rekonstruieren und zu analysieren. Das Datenmaterial, das in zwei Datenhandbüchern⁴ dokumentiert ist, umfasst die historische Studentenstatistik in Form von Langzeitreihen. Im ersten Teilband sind die aggregierten Daten für den Hochschulbesuch auf gesamtstaatlicher Ebene dokumentiert. Über die Universitäten hinaus sind vor allem auch die Technischen Hochschulen einbezogen worden. Im zweiten Teilband sind die Daten für den langfristigen Universitätsbesuch in Deutschland auf der Ebene der einzelnen Hochschulen dokumentiert.

Im Rahmen eines vom Bundesministerium für Bildung und Forschung geförderten Projektes, das sich speziell mit historischen Innovationsindikatoren beschäftigte, wurden die Daten in eine MS Access-Datenbank überführt.⁵ Die Datenbank stellt eine digitale Kopie der zwei Datenhandbücher dar. Aus der Datenbank können die Daten in zwei verschiedenen Arten ausgewählt werden: Erstens als vollständige oder Teiltabelle mit denselben Daten, wie sie in einem der beiden Datenhandbücher zu finden sind und zweitens als Tabelle, deren Zeitreihen vom Benutzer selbst zusammengestellt werden können. Auf Einzelheiten sowie die entsprechenden Ergebnisse dieses Vorgehens wird im Abschnitt IV noch eingegangen. Es sei an dieser Stelle aber ausdrücklich betont, dass eine Entschlüsselung der auf FORTRAN basierenden, heute eher kryptisch anmutenden Datenformate, in dem die Daten 1995 an das ZHSF abgegeben wurden, ohne die Unterstützung durch einen der Primärforscher (Volker Müller-Benedict, Göttingen), kaum oder mit einem vertretbaren Arbeitsaufwand nicht möglich gewesen wäre. Dies sei hier deshalb erwähnt, weil dieser Umstand für den künftigen Umgang mit älteren Datenbeständen der „Historischen Statistik“ ein wohl generelles Problem darstellen dürfte.

Nachdem die Daten in eine Access-Datenbank überführt waren, lag es nahe, diese auch online verfügbar zu machen, um so ihre direkte Nutzung über das Internet zu ermöglichen. Das Ergebnis dieser Bemühungen ist die Internet-Datenbank HISTAT (www.histat.gesis.org), die ebenfalls im nächsten Abschnitt ausführlich dargestellt wird, so dass hier einige grundsätzliche Anmerkungen genügen mögen. Die Datenbank ist themen- und studienorientiert auf-

⁴ Titze, H. unter Mitarbeit von Herrlitz, H.-G.; Müller-Benedict, V.; Nath, A.: Datenhandbuch zur Deutschen Bildungsgeschichte. Bd. Titze, H. unter Mitarbeit von Herrlitz, H.-G.; Müller-Benedict, V.; Nath, A.: Datenhandbuch zur Deutschen Bildungsgeschichte. Bd. 1: Hochschulen. Teil 1: Das Hochschulstudium in Preußen und Deutschland 1820-1944. Göttingen 1987 sowie Titze, H. unter Mitarbeit von Herrlitz, H.-G.; Müller-Benedict, V.; Nath, A.: Datenhandbuch zur Deutschen Bildungsgeschichte. Bd. Titze, H. unter Mitarbeit von Herrlitz, H.-G.; Müller-Benedict, V.; Nath, A.: Datenhandbuch zur Deutschen Bildungsgeschichte. Bd. 1: Hochschulen. Teil 2: Wachstum und Differenzierung der deutschen Universitäten 1830-1945. Göttingen 1995.

⁵ Metz, Rainer; Watteler, Oliver: Historische Innovationsindikatoren. In: Historical Social Research/Historische Sozialforschung, 27/1, 2002, S. 4-129.

gebaut. D.h. es gibt verschiedene Themen,⁶ denen einzelne Studien zugeordnet werden. Die Vorteile einer solchen Konzeption – auf die Nachteile wird gleich noch einzugehen sein – liegen auf der Hand. So lässt sich das Spektrum der Themen, unter denen die einzelnen Studien eingeordnet werden, beliebig erweitern. Auch können jederzeit weitere Studien in die Datenbank aufgenommen werden. Liegen die Daten einmal in einem bestimmten Format vor, lassen sie sich relativ einfach in die Datenbank integrieren. Damit bietet die Datenbank eine ideale Möglichkeit, die Daten der „Historischen Statistik“ mit einem relativ geringen Arbeits- und Zeitaufwand online verfügbar und recherchierbar zu machen.⁷

Allerdings sind mit dieser Datenbankkonzeption auch Nachteile verbunden. Das Ordnungssystem der Datenbank sind Themen und Studien, nicht aber eine Systematik der Variablen (wie z.B. im Statistischen Jahrbuch). Damit ermöglicht die Datenbank eine Recherche nur innerhalb der jeweils ausgewählten Studie. Interessiert sich der Nutzer z.B. für Sterberaten in Deutschland, muss er zunächst eine Studie auswählen, und erst dann kann er nach der ihn interessierenden Variablen suchen. Die Auswahl wird ihm durch die vorgenommene Zuordnung der Studie zu einem bestimmten Thema erleichtert. Will er einen Gesamtüberblick über alle Ausprägungen der ihn interessierenden Variablen erhalten, muss er alle relevanten Studien durchsuchen. Dieser Zustand ist insofern unbefriedigend, als eine solche Recherche voraussetzt, dass der Nutzer die für ihn relevanten Studien kennt. Das Ergebnis der Recherche wird damit entscheidend davon abhängen, welche Studie der Nutzer auswählt.

Wünschenswert ist deshalb eine Datenbank, die die Recherche systematisch über die einzelnen Studien hinweg ermöglicht. Voraussetzung dafür ist eine Datenbank, die die Variablen systematisch aufführt, unabhängig davon, in welcher Studie sie vorkommen. Nur somit ist für den Nutzer die Möglichkeit gegeben, z.B. nach der Variablen „Sterberate“ zu suchen, unabhängig davon, in welcher Studie diese Variable vorkommt. Dieser bislang noch nicht realisierte Schritt erfordert zunächst die Festlegung einer Systematik und sodann die Klassifikation der einzelnen in den Studien enthaltenen Variablen. Die Konzeption einer solchen Datenbank und die Aufbereitung der gegenwärtig in den verschiedenen Studien enthaltenen Variablen ist in Arbeit. Ihre Realisierung ist ein wichtiger Schritt hin zu einem elektronischen Handbuch zur „Historischen Statistik“ von Deutschland, das die Bereiche Bevölkerung, Wirtschaft, Gesellschaft und Staat regional tief gegliedert und intertemporal vergleichbar erfasst.

⁶ Wie z.B.: Demographie, Bildung und Wissenschaft, Geld und Kredit, Innovation, Kommunikation und Verkehr, Gesundheitswesen, Preise, Sozialer Wandel, Staat und öffentliche Ausgaben, Unternehmen, Wachstum und Konjunktur.

⁷ Selbstverständlich werden nur solche Daten online angeboten, die der Primärforscher zur Nutzung freigegeben hat. Auch wird darauf zu achten sein, dass bei der Verwendung der Daten durch Dritte der Primärforscher und nicht das ZHSF als Urheber genannt wird.

IV

Am Beispiel der Daten aus dem bereits erwähnten Verbundprojekt „Deutsche Bildungsstatistik 1800-1945“⁴ (im Folgenden „QUAKRI-Daten“ genannt) sollen nun die einzelnen Arbeitsschritte erläutert werden, die erforderlich waren, um die Rohdaten zunächst in die MS Access-Datenbank (im Folgenden „Quakri-Datenbank“ genannt) und sodann in die Internet-Datenbank HISTAT umzusetzen. An die Datenbanken wurden dabei folgende Anforderungen gestellt:

- 1) die Daten/Zeitreihen müssen sich direkt in Formate für die statistische Auswertung (Excel, SPSS, etc.) exportieren lassen,
- 2) die Datenbank muss die Eingabe neuer Zeitreihen ermöglichen,
- 3) die Datenbank muss die Integration neuer Projekte/Studien erlauben,
- 4) die Datenbank muss eine detaillierte Kommentierung einzelner Datenpunkte einer Zeitreihe erlauben, zudem muss
- 5) die Internet-Datenbank die Übernahme der Daten eines Desktop-Datenbankmodells ermöglichen.

Datenkonvertierung. Die QUAKRI-Daten umfassen etwa 56.000 Zeitreihen, die ursprünglich in verschlüsselter und stark komprimierter Form auf 136 ASCII-Dateien verteilt waren. Die Daten wurden mit einem FORTRAN-Programm erfasst,⁸ mit dem seinerzeit sämtliche Auswertungen und Grafiken für die beiden bereits erwähnten Datenhandbücher zur Deutschen Bildungsgeschichte generiert wurden.

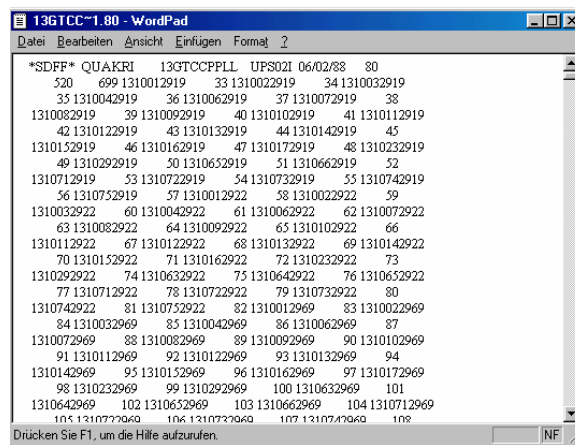


Abb. 1: Beispiel für „QUAKRI-Daten“ in komprimierter Form

⁸ Entwickelt von V. Müller-Benedict (Göttingen, Methodenzentrum Sozialwissenschaften)

Aufgrund der zu jener Zeit knappen Speicherkapazität wurde ein 10stelliger Schlüssel konzipiert, der sämtliche Informationen zu den Zeitreihen enthält. Z.B. ist die Zeitreihe für die weiblichen Studentinnen aus Preußen, welche an der Göttinger Universität im Fach Biologie immatrikuliert waren, mit folgendem Schlüssel versehen: 0121270400. Dabei bedeuten: 01 - Studentenzahl, 2 - weiblichen Geschlechts, 1 - aus Preußen, 27 - Studium der Biologie/dito einschl. Botanik, 04 - in Göttingen, 00:- immatrikuliert. Das Schlüsselssystem wurde für die QUAKRI-Datenbank übernommen. In einem ersten Schritt wurden nun diese Daten mit einem C⁺⁺-Script in ein von MS-Access lesbares Format transformiert, da es eine schnelle und effektive Zeitreihenauswahl auch bei großen Datenmengen erlaubt.

QUAKRI-Datenbank. In MS-Access wurde dann eine Desktop-Datenbank entwickelt, die den ersten beiden oben genannten Anforderungen genügt. In die Datenbank wurden sämtliche QUAKRI-Daten (welche vor dem Import noch einer Plausibilitätsprüfung unterzogen wurden) übernommen, sie ist intuitiv bedienbar, ermöglicht die freie Suche nach Zeitreihen, einen direkten Export der Daten sowie die Auswahl von Tabellen, entsprechend der beiden Datenhandbücher.

The screenshot shows the 'QUAKRI 1.1-2000 - [Zentralarchiv für Empirische Sozialforschung]' window. The main title is 'Zeitreihenauswahl'. Below it, there's a section 'Die Zeitreihe enthält Daten aus folgendem Bereich:' with buttons for 'Fächer/Fakultäten', 'Semesterbelegung', 'Studenten', 'Väterberufe', 'Fakultätswechsel', and 'Studienför'. The 'Fächer/Fakultäten' button is selected.

Below this is a section 'Die absoluten Zahlen der Studenten nach Fakultäte' with input fields for 'Geschlecht' (Gesamtzahl ohne Trennung), 'Regionale Herkunft' (Preußen), 'Studienfach' (Evangelische Theologie), 'Universität' (Berlin), and 'Zusatzindikator' (Immatrikulierte Studenten insg. (Uni)). There are buttons for 'bestätigen', 'Nächste Einstellung', and 'Vorherige Ein'.

At the bottom right is a button 'Ansicht/Export'.

On the right side, there's a table titled 'QUAKRI 1.1-2000 - QUAKRI3 : Kreuztabellenabruf'. The table has columns: 'Jahr_Sem', '1', '2', '3', and 'Die'. The rows show data for various years and semesters, with values ranging from 15870 to 12435.

Jahr_Sem	1	2	3	Die
1830 WS	15870	179	16049	
1831 SS	15466			
1831 WS	14412	284	14696	
1832 SS	14186			
1832 WS	14165	498	14663	
1833 SS	13728			
1833 WS	13384	594	13978	
1834 SS	12896			
1834 WS	12575	594	13169	
1835 SS	11929			
1835 WS	11930	575	12505	
1836 SS	11566			
1836 WS	11564	607	12171	
1837 SS	11492			
1837 WS	11725	601	12326	
1838 SS	11674			
1838 WS	11815	639	12454	
1839 SS	11555			
1839 WS	11761	674	12435	

At the bottom, there's a status bar showing 'Datensatz: 1 von 233'.

Abb. 2: Beispiel für ein Auswahl-Formular und ein Auswahl-Ergebnis der Desktop-Datenbank

Die dritte und vierte oben genannte Anforderung an die QUAKRI-Datenbank, also die Integration neuer Studien sowie die Möglichkeit einer detaillierten Kommentierung einzelner Datenpunkte, die i.d.R. bereits bei der Erhebung der Daten notwendig wird, wurde inzwischen im Rahmen eines DFG-Projektes an der Universität Göttingen realisiert.⁹ Diese weiter entwickelte QUAKRI-Datenbank lässt sich mit relativ geringem Programmieraufwand an jedes Projekt, welches die Erhebung, Darstellung und Auswertung von Daten der „Historischen Statistik“ zum Inhalt hat, anpassen. Von besonderem Vorteil ist hierbei der Umstand, dass sich die Daten nach Abschluss des Projekts direkt in die Internet-Datenbank HISTAT portieren lassen. Es muss aber betont werden, dass sich in die HISTAT-Datenbank auch Zeitreihen bzw. Studien integrieren lassen, die mit anderen Programmen/Datenbanken erfasst wurden.

HISTAT-Datenbank. Zur Realisation der Internet-Datenbank HISTAT (www.histat.gesis.org) wurde das Datenbank-Modell (Entity Relationship Modell, ERM) der QUAKRI-Datenbank benutzt. Die Internet-Datenbank wurde auf der technischen Grundlage eines LAMP-Systems (Linux als Betriebssystem, Apache als Webserver, MySQL als Datenbank und PHP als Programmiersprache) entwickelt. Die Datenbank ist, wie bereits erwähnt, themen- und studienorientiert (vgl. Fußnote 6) aufgebaut. Die QUAKRI-Daten sind dabei unter dem Thema „Bildung und Wissenschaft“ in der Datenbank zu finden. Die Datenbank enthält nicht nur die Zeitreihen verschiedener Studien, sondern auch die durch das ZHSF erstellten Studienbeschreibungen bzw. Kontextinformationen. Die Datenbank erlaubt die Zusammenstellung von Zeitreihen unterschiedlicher Studien sowie den Download der Abfrageergebnisse als Excel-Dateien oder als csv-Dateien. Damit lassen sich die recherchierten Daten direkt in eigene Studien einbeziehen und unter neuen Gesichtspunkten auswerten. Trotz Einsatz unterschiedlicher Technologien ist es damit gelungen, Oberflächen und Navigation der QUAKRI- und HISTAT-Datenbank ähnlich zu gestalten, was die Bedienung erleichtert.

Auch der Wunsch, bereits am ZHSF archivierte Studien zur „Historischen Statistik“ in die Internet-Datenbank zu importieren, konnte realisiert werden. Ohne größeren Aufwand ist dies möglich, sofern die Daten als Excel-Dateien oder in nach Excel importierbare Formate (1-2-3 Lotus, SPSS, csv) vorliegen. Ein VB.net-Modul übernimmt den direkten Import dieser Daten in die MySQL-Datenbank. Der Arbeitsablauf ist in Abbildung 3 visualisiert.

⁹ Durch Müller-Benedict und den Zweitautor.

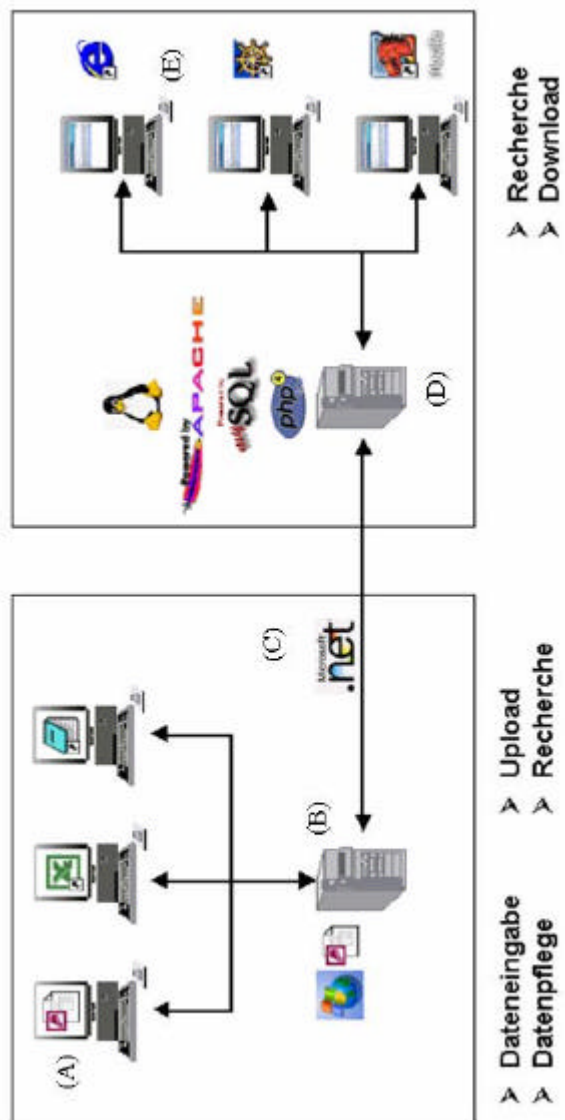


Abb. 3: Arbeitsablauf beim Import von ZHSF-Studien in die Internet-Datenbank (A -D) und deren Darstellung im Internet (E)

Die Studien der „Historischen Statistik“ werden am ZHSF aufbereitet (A) und auf dem Netzwerkservers gespeichert (B). Ein VB.net-Modul (C) transformiert die Daten in das Format der MySQL-Datenbank- und überträgt die Daten via ODBC auf dem Linux-Server (D). Bei diesem Schritt wird für jede Zeitreihe ein eindeutiger Schlüssel generiert (s.o.). Mittels eines Browsers (E) kann jeder Nutzer in der Datenbank recherchieren und die ausgewählten Daten downloaden.

Modularer Ansatz. Von der Aufbereitung der QUAKRI-Daten bis zu ihrer Darstellung im Internet wurde ein modularer Ansatz verfolgt. Grundlage hierfür ist ein einheitliches Datenmodell (ERM), welches in den verschiedenen Teilprojekten mit ihrer jeweiligen Software zur Datenverarbeitung Verwendung findet. So wurde, wie bereits erwähnt, die QUAKRI-Datenbank am Methodenzentrum Sozialwissenschaften der Universität Göttingen um zusätzliche Module erweitert, die u.a. eine exakte Dokumentation (mit integrierter Literatur-Datenbank) sowie weitere Eingabeoptionen beinhalten.

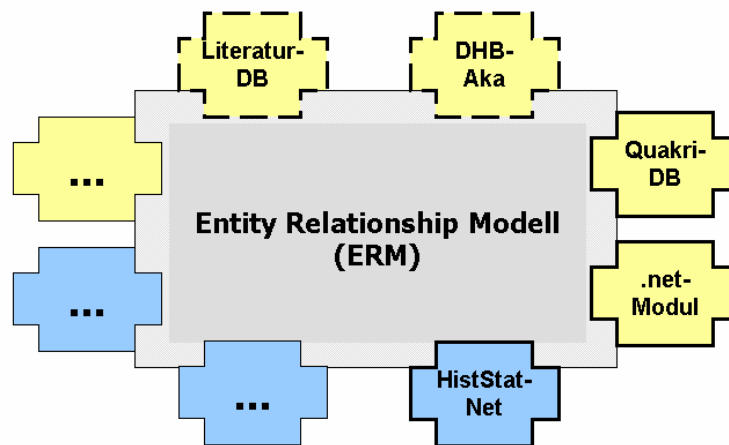


Abb. 4 : Modularer Ansatz bei der Aufbereitung und Ausgabe von Daten zur Historischen Statistik (gelb: Desktop-Module; blau: Internet-Module); Punkte stehen für die Erweiterbarkeit

Dieser Ansatz erlaubt es, sowohl die Desktop- wie auch die Internet-Datenbank sukzessive weiter zu entwickeln. Zurzeit sind zwei Erweiterungen der HISTAT-Datenbank geplant bzw. in Arbeit: 1. die Darstellung von Anmerkungen und Quellen zu jedem einzelnen Datenpunkt und 2. eine Studienübergreifende Metasuche.

Open-Source-Software. Wenngleich die QUAKRI-Datenbank auf der Basis einer proprietären Software (MS-Access) entwickelt wurde, so stehen doch alle entwickelten Module unter der Open-Source-Lizenz „Gnu Public License“ (GPL). Die Internet-Datenbank „HISTAT“ nutzt ausschließlich Open-Source-Software als Entwicklungswerkzeuge (Linux, Apache, MySQL, PHP) und steht ebenfalls unter der GPL. Damit wird ein sicherer Weg zur langfristigen Archivierung digitaler Daten der „Historischen Statistik“ am ZHSF beschritten.